

Un algoritmo para transformar en variables métricas respuestas de preferencia

Vicente Manzano – 2005

Situación

A un conjunto de n sujetos se les presenta el mismo conjunto de s estímulos. De ellos han de escoger individualmente k , ordenados por preferencia. De este modo, se genera una matriz de dimensiones $n \times k$.

Objetivo

Se pretende asignar a cada uno de los s atributos una puntuación métrica, p , comprendida entre los valores min y max . [$p \in (min, max)$].

Unidad básica

Las respuestas de preferencia pueden ser codificadas de modo muy distinto. Usualmente se asignan los primeros k números naturales, de tal forma que el estímulo escogido en primer lugar recibe la codificación 1, el segundo el valor 2, y así sucesivamente. En tal caso, para que sea considerada una medida de preferencia, es necesario darle la vuelta a la escala, de tal forma que un número mayor implique, a su vez, mayor preferencia. Por otro lado, para unificar criterios de cálculo, es conveniente convertir la escala en el intervalo $(0,1)$, donde 0 implica que el estímulo no ha sido seleccionado, mientras que 1 indica que el estímulo se ha seleccionado en primer lugar (es el más preferido). Si la puntuación original de preferencia que el sujeto j ha asignado al atributo i es simbolizada con g_{ij} , y la puntuación transformada en $(0,1)$ con h_{ij} , entonces:

$$h_{ij} = \frac{k - g_{ij} + 1}{k} = 1 - \frac{g_{ij} - 1}{k} \quad (1)$$

La siguiente tabla muestra diferentes valores h_{ij} , en función de k y de g_{ij} .

h_{ij}		k							
		2	3	4	5	6	7	8	9
g_{ij}	1	1	1	1	1	1	1	1	1
	2	0,500	0,667	0,750	0,800	0,833	0,857	0,875	0,889
	3	-	0,333	0,500	0,600	0,667	0,714	0,750	0,778
	4	-	-	0,250	0,400	0,500	0,571	0,625	0,667
	5	-	-	-	0,200	0,333	0,429	0,500	0,556
	6	-	-	-	-	0,167	0,286	0,375	0,444
	7	-	-	-	-	-	0,143	0,250	0,333
	8	-	-	-	-	-	-	0,125	0,222
	9	-	-	-	-	-	-	-	0,111

Procedimiento 1

El mejor de los atributos o estímulos posible es aquél que es escogido en primer lugar por los n sujetos, lo que implicaría una puntuación máxima de valor n . En el otro extremo, se encuentra el atributo o estímulo jamás escogido, cuya puntuación será, por tanto, 0. El primer procedimiento, el más inmediato, para obtener una puntuación final para cada atributo será expresar la suma de las n puntuaciones h_{ij} obtenidas entre n , lo que expresa una puntuación media que se sitúa en el intervalo (0,1):

$$p_i^{(1)} = \frac{1}{n} \sum_{j=1}^n h_{ij}$$

No obstante, este índice tiene algunos inconvenientes. En primer lugar, cabe esperar que los valores obtenidos para la puntuación final sean más bien bajos, puesto que es harto difícil que un listado de s atributos no permita cierta repartición de las preferencias. Por otro lado, el índice no es sensible al número s de atributos, lo que no es deseable. Cabe concebir a un atributo como mucho mejor o más meritorio cuanto mayor sea el número de alternativas de elección que tienen los sujetos a su disposición. Escoger un estímulo entre dos posibilidades no es gran cosa. Escogerlo sistemáticamente entre un conjunto de una docena, ya es suficiente. Para solucionar este inconveniente, se propone el procedimiento 2.

Procedimiento 2

Los valores $p_i^{(1)}$ se encuentran comprendidos en el intervalo (0,1), por lo que una potencia inferior a la unidad los eleva de cuantía. Una estrategia para incrementar su valor en el mismo intervalo, aumentado la capacidad de discriminar entre puntuaciones, a la vez que se tienen en cuenta los valores k y s es utilizar la potencia k/s . Conforme mayor sea el número s de atributos de entre los que se ha de escoger preferencias, la potencia será menor y, por tanto, mayor será el resultado al aplicársela a la base $p_i^{(1)}$. Conforme mayor sea el número de atributos k a seleccionar, menos mérito tendrán las puntuaciones por atributo y así quedará registrado en el nuevo índice, puesto que aumentará el valor de la potencia y el resultado será menor. Así pues, se propone:

$$p_i^{(2)} = (p_i^{(1)})^{\frac{k}{s}} = \left(\frac{1}{n} \sum_{j=1}^n h_{ij} \right)^{\frac{k}{s}}$$

Este índice tiene, no obstante, un inconveniente de interés. Los valores de $p_i^{(2)}$ se sitúan en el intervalo (0,1) sin expresar ningún sentido. Sería deseable que el índice pudiera mostrar en qué medida las preferencias obtenidas se encuentran por debajo o por encima de lo que cabría esperar por azar o, al menos, en relación a una preferencia difusa. Esto es lo que se pretende remediar con el procedimiento 3.

Procedimiento 3

Cada sujeto entrevistado suministrará un total de t puntos, donde:

$$\begin{aligned} t &= \sum_{i=1}^k h_{ij} = \sum_{i=1}^k \left(1 - \frac{i-1}{k}\right) = k - \frac{1}{k} \left(\sum_{i=1}^k i - k\right) = \\ &= k - \frac{1}{k} \left(\frac{k(k+1)}{2} - k\right) = k - \frac{k+1}{2} + 1 = \frac{k+1}{2} \end{aligned}$$

Para toda la muestra, se recibirán tn puntos que, bajo el supuesto de que todos los atributos resultaran igualmente atractivos, implicarían un valor esperado, v , de cuantía:

$$v = \frac{tn}{s} = \frac{n}{s} \frac{k+1}{2}$$

Una forma de expresar en qué medida las preferencias mostradas por un atributo se encuentran por encima o por debajo de la preferencia esperada, es:

$$p_i^{(3)} = \frac{\sum_{j=1}^n h_{ij} - v}{n - v}$$

o, incluyendo las modificaciones del procedimiento 2:

$$p_i^{(4)} = \left(\frac{\sum_{j=1}^n h_{ij} - v}{n - v} \right)^{\frac{k}{s}}$$

Un programa sencillo BASIC

El archivo de disco que se lee cuenta con un registro inicial con tres informaciones ordenadas: número de atributos, número de selecciones y tamaño de la muestra. Sigue una matriz de $s \times k$: s filas, cada una para un atributo, y k columnas, donde figura el número de veces que ese atributo ha sido seleccionado en primer, segundo, tercer..., k -ésimo lugar.

```
10 CLS : KEY OFF
20 REM
30 REM Lee y escribe los datos
40 REM
50 OPEN "datos" FOR INPUT AS #1
60 INPUT #1, S, K, N
70 V=(K+1)/S/2*N
90 OPEN "salida" FOR OUTPUT AS #2
100 REM
110 REM Escribe presentación en la salida
```

```

120 REM
130 PRINT #2, "Archivo de entrada: 'datos'" : PRINT #2,
140 PRINT #2, "      ";S;"atributos."
150 PRINT #2, "      ";K;"elementos de selección."
160 PRINT #2, "      ";N;"sujetos entrevistados." : PRINT #2,
170 PRINT #2, "Matriz de datos:" : PRINT #2,
180 FOR A=1 TO S
190   PRINT #2, "  Atributo"; : PRINT #2, USING "##"; A;
200   MAT=0
210   FOR B=1 TO K
220     INPUT #1, NUM
230     PRINT #2, USING "#####";NUM;
240     MAT=MAT+NUM*(1-(B-1)/K)
250   NEXT B
260   H=MAT/N
270   PUN=(MAT-V)/(N-V)
280   IF PUN>0 THEN PAN=PUN^(K/S) ELSE PAN=-(-PUN)^(K/S)
290   PRINT #2, USING "####.###";MAT;MAT-V;H;H^(K/S);
300   PRINT #2, USING "####.###";PUN;PAN
310 NEXT A
320 PRINT #2, : PRINT #2, "  Valor esperado por atributo:";
330 PRINT #2, USING "#####.##";VE
340 CLOSE : SYSTEM

```

Un ejemplo

Archivo de entrada:

```

9 3 1870
1163 336 157
320 384 373
146 213 270
134 198 209
70 139 187
45 107 165
40 407 199
20 78 162
8 32 48

```

Archivo de salida:

Archivo de entrada: 'datos'

```

9 atributos.
3 elementos de selección.
1870 sujetos entrevistados.

```

Matriz de datos:

Atributo 1	1163	336	157	1439.333	1023.778	0.770	0.916	0.704	0.890
Atributo 2	320	384	373	700.333	284.778	0.375	0.721	0.196	0.581
Atributo 3	146	213	270	378.000	-37.556	0.202	0.587	-0.026	-0.296
Atributo 4	134	198	209	335.667	-79.889	0.180	0.564	-0.055	-0.380
Atributo 5	70	139	187	225.000	-190.556	0.120	0.494	-0.131	-0.508
Atributo 6	45	107	165	171.333	-244.222	0.092	0.451	-0.168	-0.552
Atributo 7	40	407	199	377.667	-37.889	0.202	0.587	-0.026	-0.296
Atributo 8	20	78	162	126.000	-289.556	0.067	0.407	-0.199	-0.584
Atributo 9	8	32	48	45.333	-370.222	0.024	0.289	-0.255	-0.634

Valor esperado por atributo: 415.56